

---

## Présentation : « Les » linguistiques de corpus : enjeux spécifiques inhérents à la nature des données traitées

Sophie Azzopardi

---



### Édition électronique

URL : <http://journals.openedition.org/praxematique/1169>

DOI : 10.4000/praxematique.1169

ISSN : 2111-5044

### Éditeur

Presses universitaires de la Méditerranée

### Édition imprimée

Date de publication : 1 janvier 2010

Pagination : 209-220

ISBN : 978-2-36781-012-6

ISSN : 0765-4944

### Référence électronique

Sophie Azzopardi, « Présentation : « Les » linguistiques de corpus : enjeux spécifiques inhérents à la nature des données traitées », *Cahiers de praxématique* [En ligne], 54-55 | 2010, document 12, mis en ligne le 01 janvier 2013, consulté le 23 septembre 2020. URL : <http://journals.openedition.org/praxematique/1169> ; DOI : <https://doi.org/10.4000/praxematique.1169>

---

Tous droits réservés

## **Présentation : « Les » linguistiques de corpus : enjeux spécifiques inhérents à la nature des données traitées**

Ce numéro double entend placer les corpus au centre de la réflexion. La problématique n'est pas nouvelle, et cette place centrale est celle que « le » corpus en tant que notion comme « les » corpus en tant qu'objets occupent dans les recherches de plusieurs domaines depuis plusieurs dizaines d'années. En 2005, Damon Mayaffre met en évidence la prépondérance des corpus en linguistique, à une époque où ils deviennent un enjeu institutionnel et financier :

Le corpus — la notion et l'objet — risque d'être victime aujourd'hui en France de son succès. Plus une discipline, plus un comité scientifique, plus un chercheur qui n'y fasse référence ; plus un linguiste, surtout, qui ne le manipule, le caresse ou le maltraite.

(Mayaffre 2005 : 5)

Le corpus est donc un incontournable des analyses linguistiques, à tel point que de nombreuses manifestations et un nombre tout aussi important de publications ont pour objet de réflexion le corpus et les enjeux de son utilisation au sein des sciences du langage. Une revue a émergé de la nécessité de mener une réflexion constante autour de cette notion (*Corpus* 2002 et *sq.*), des recueils (Williams 2005) et des numéros thématiques analysent la place et dressent déjà le bilan de l'utilisation des corpus en linguistique (*Revue française de linguistique appliquée* 2007/1, *Langages* 2008/3). On est pourtant loin d'avoir fait le tour de la question comme le prouve la sortie très récente du numéro de *Langages* consacré à l'analyse de corpus (*Langages* 2012/3). C'est pourquoi cette présentation entend donner un aperçu des problématiques posées par le recours au(x) corpus dans les analyses linguistiques.

Après un premier numéro consacré aux questions et enjeux qui traversent l'ensemble des recherches menées dans le cadre d'une approche essentiellement *corpus-driven* (au cours duquel ont été abordées entre autres les questions de la représentativité du corpus, de la place du chercheur au sein de l'analyse et de l'exploitation pluridisciplinaire des corpus), ce second numéro se donne pour objectif de faire émerger les spécificités inhérentes aux différents types de corpus analysés afin de rendre compte de la présence de plusieurs linguistiques de corpus comme l'ont fait précédemment par exemple Benoît Habert, Adeline Nazarenko et André Salem (1997). Il est cependant intéressant de voir que quinze ans après la publication de cet ouvrage intitulé *Les linguistiques de corpus*, notre emploi de ce pluriel et le choix de parler « des » linguistiques de corpus ne se fonde pas sur les mêmes constats. Habert, Nazarenko & Salem envisagent en effet la diversité des pratiques de constitution de corpus écrits uniquement puisqu'à ce moment-là, les corpus accessibles étaient principalement des corpus textuels :

Nous avons mis l'accent sur les corpus relevant de l'écrit. Les corpus d'oral transcrit sont encore rares : la transcription proprement dite, les choix qu'elle entraîne, les coûts qu'elle suppose freinent leur développement, même si celui-ci semble s'accélérer dans les dernières années.  
(1997 : 7)

Nous avons emprunté ce terme pluriel ici pour désigner un ensemble plus diversifié encore de linguistiques de corpus en fonction de la nature des corpus étudiés. On distinguera dans ce numéro trois types de pratiques correspondant à trois natures de corpus : celles portant sur des corpus textuels (textes écrits), celles portant sur des corpus vidéo, et enfin celles portant sur des corpus multimodaux <sup>1</sup>.

---

1. On entend par *corpus vidéo* ceux qui proviennent d'enregistrements vidéo, sans que ceux-ci soient nécessairement multimodaux (le chercheur peut ne vouloir constituer qu'un corpus visuel à partir de ces enregistrement). Les *corpus multimodaux* sont ceux qui sont constitués en vue d'analyser plusieurs modalités (gestes, paroles) et se situant sur plusieurs domaines d'analyse (phonétique, prosodie, syntaxe, morphologie...).

## **I. Après le déploiement de la linguistique de corpus, le foisonnement des linguistiques de corpus**

La linguistique de corpus, caractérisée par une approche *bottom-up*, trouve son origine au sein des perspectives de recherche envisagées dans les linguistique du Traitement automatique du langage (TAL). Sa généralisation à l'ensemble des analyses linguistiques et son élargissement à d'autres disciplines comme la recherche clinique ou la sociologie est en revanche beaucoup plus récent. Depuis la fin des années 1990 en effet, l'essor et la démocratisation d'Internet a permis un accès beaucoup plus aisé aux données. Parallèlement à cela, les outils permettant de transcrire et d'exploiter les corpus constitués se sont eux aussi développés, favorisant le déploiement du recours au corpus en recherche fondamentale comme en recherche appliquée.

La remarquable avancée technologique de ces dernières années est donc ce qui selon nous, se trouve à l'origine non seulement de la généralisation du recours au corpus dans les pratiques des chercheurs en linguistique, mais aussi de la diversification des corpus constitués. À l'heure actuelle, le chercheur peut en effet avoir accès à des dispositifs d'enregistrement audio et/ou vidéo facilement manipulables et transportables. Les logiciels de traitement du son, de l'image et de la parole sont de plus en plus perfectionnés, et des sessions de formation sont mises en place au sein des centres de recherche pour que chaque chercheur puisse bénéficier de ces nouveaux outils en constante évolution. Il n'est plus nécessaire d'avoir une formation avancée en informatique pour disposer de tels moyens. De ce fait, la constitution de corpus autres que textuels suscite un intérêt croissant et l'exploration de ces nouvelles données élargit les perspectives d'analyse et le champ des phénomènes étudiés.

Si les problèmes posés d'une part par les choix opérés lors de la constitution de corpus et d'autre part par la place de l'intuition du chercheur lors de la collecte, de la transcription et de l'exploitation des données sont communes à toutes les linguistiques de corpus, les stratégies mises en place et les moyens utilisés sont spécifiques à chaque type de corpus. On pense par exemple à la prise en compte de la temporalité et de la spatialité des réalisations discursives. Le problème ne se pose pas dans des termes similaires selon que l'on a affaire à un corpus textuel s'inscrivant dans une certaine linéarité ou que l'on doive analyser des

corpus multimodaux. Lorsque plusieurs phénomènes sont observables au même moment, on peut se demander comment rendre compte de chacun d'eux au sein de l'analyse alors qu'elle possède elle-même sa propre chronologie. Les corpus multimodaux, comme les corpus vidéo, constituent un ensemble de données hétérogènes qui ne peuvent être étudiées au moyen des mêmes méthodes que celles utilisées pour les corpus textuels qui ne donnent à voir qu'un seul type de donnée.

Ces nouveaux types de corpus impliquent que l'on revienne sur la question de la place du chercheur et de la subjectivité qu'il introduit nécessairement dans les choix qui président à la constitution des corpus comme dans les résultats auxquels il parvient. Les corpus textuels présentent l'avantage de constituer des données « brutes », authentiques. Le chercheur semble dans ce cas mener une réelle approche *bottom-up* en ce sens qu'il observe (presque) directement les réalisations telles qu'elles se présentent à l'origine. Dans le cas de corpus vidéo ou multimodaux, le dispositif d'enregistrement s'impose comme un intermédiaire entre les données et le chercheur et impose des contraintes techniques et matérielles qui influencent la façon dont le corpus va être constitué. Il existe en effet un décalage entre ce qui est observé par l'œil humain et ce qui est capté par un dispositif d'enregistrement. Dans le cadre d'une approche *corpus-driven*, cette particularité des corpus vidéo et multimodaux doit être prise en compte et faire l'objet d'une réflexion.

Les linguistiques de corpus, en s'opposant aux linguistiques traditionnelles, prônent une démarche de recherche à partir des usages discursifs en constante évolution qui va à l'encontre d'une conception de la langue comme système figé. Or, on s'aperçoit que cet intérêt pour les usages en linguistique va de pair avec une attention portée aux usagers de la langue. On ne s'étonnera donc pas, comme le souligne Wolfgang Teubert (2009), qu'un certain nombre de travaux de recherche conduits à partir de corpus se situent à la frontière de la linguistique et de la sociologie. Cela semble d'autant plus vrai avec le développement de corpus vidéo et multimodaux qui permettent de rendre compte d'un certain nombre de phénomènes co-verbaux analysables en termes sociolinguistiques que les corpus textuels ne sont pas à même de retranscrire.

Les travaux rassemblés dans ce numéro permettent de donner un aperçu des spécificités inhérentes à chaque type de corpus (textuels,

vidéo, multimodaux) aux cours des différentes étapes par lesquelles passe le chercheur dans son travail de recherche. De la constitution à l'exploitation des corpus en passant par leur transcription, il a paru intéressant au vu des contributions de faire émerger ces particularités au sein de la réflexion après avoir présenté ce qui les rassemblait. On procèdera pour cela en trois temps. Dans un premier temps, les articles portant sur des corpus textuels. Bien que les questions que suscitent ces corpus ne soient pas nouvelles, il nous a paru important que ce type de corpus autrefois prédominant ne soit pas oublié. L'évolution des technologies comme celle des perspectives de recherche et de leurs visées applicatives soulèvent de nouvelles questions et créent de nouveaux enjeux auxquels il nous semble devoir réfléchir. Les travaux présentés en deuxième partie sont consacrés quant à eux aux corpus vidéo. Leurs objectifs d'analyse sont très différents, mais on s'aperçoit que les contraintes techniques imposées par ce type de dispositif d'enregistrement soulèvent des questions similaires. Enfin, nous avons souhaité clore ce numéro, et par conséquent la réflexion menée dans les deux numéros autour des corpus, par des articles abordant les aspects problématiques de la transcription et de l'exploitation de corpus oraux ou multimodaux. Seront traitées les questions de la temporalité, de la gestualité et de l'oralité inhérentes à la nature des données collectées puis analysées.

## **2. Problématiques actuelles et multidisciplinaires autour de l'exploitation de corpus textuels**

Si les corpus textuels constituent l'objet d'analyse le plus ancien des linguistiques de corpus, la révolution numérique de ces dernières années a considérablement modifié notre approche des textes tout comme nos pratiques de lecture et d'écriture. Damon Mayaffre propose de réfléchir aux relations entre corpus textuels et web-corpus afin de mettre en évidence les changements qui interviennent au niveau des pratiques des chercheurs face à ces nouveaux types de textes. L'auteur pose ainsi la nécessité de redéfinir la notion de corpus textuel, notamment en introduisant une différenciation entre corpus papier et corpus numérique. Au-delà de cette problématique, on a vu que l'essor d'Internet et des nouvelles technologies a permis un accès grandissant à un nombre de données croissant. Partant de ce constat, et de

la définition du corpus comme un ensemble construit, D. Mayaffre soulève le problème de la cohérence et de l'homogénéité du corpus. À la lumière de la révolution numérique, ce sont les délimitations et les critères définitoires du corpus textuel en tant qu'entité qui doivent être repensées.

Après cette ouverture théorique et épistémologique sur les corpus textuels à l'heure du numérique, l'article suivant met en évidence l'évolution des objectifs pour lesquels sont constitués les corpus induite par l'avancée des technologies numériques. Dans son travail, Laurence Longo présente l'étude d'un corpus multi-genres dont les visées applicatives ont pour finalité le développement d'un outil de détection automatique de thèmes permettant d'optimiser l'indexation des documents dans un moteur de recherche. La particularité de cette indexation est qu'elle prend en compte le genre du document, ce qui implique un travail sur le genre textuel. Le corpus constitué recense cinq genres textuels différents (articles journalistiques, éditoriaux, romans, lois européennes et rapports publics) que le chercheur doit caractériser en fonction de chaînes de références qui serviront de paramètres de reconnaissance pour l'indexation des documents. On voit de cette façon dans quelle mesure le numérique pose de nouvelles problématiques et demande à repenser les corpus textuels jusque dans la notion de genre discursif.

Le travail de Felice Addeo et Paolo Diana vient compléter la réflexion sur les corpus textuels en l'élargissant à d'autres disciplines. On a précédemment signalé le fait que la diversification des usages du corpus conduit inévitablement certaines analyses de corpus à la frontière de la linguistique et de la sociologie. Ce dernier article abordant les corpus textuels constitue un angle d'approche plus sociologique que linguistique et apporte un éclairage distinct sur les questions soulevées par les corpus textuels. Il s'agit ici de réfléchir aux procédés de collecte de données et de mesurer les apports et les limites de la méthode des « Histoires ». Le recours aux « Histoires » permet de collecter des informations concernant la perception que les sujets ont des « dimensions de valeur » (ex : universalisme/particularisme) en leur demandant de réagir à des narrations mises en place par le chercheur dans ce but spécifique. La question centrale ici est celle de la place du chercheur en tant qu'être subjectif non seulement en amont de la collecte, lors de l'élaboration des « Histoires », mais aussi en aval de la constitution du corpus, lors de l'interprétation des données recueillies.

Ces trois articles nous semblent avoir permis de donner un aperçu de certains des enjeux actuels et des visées applicatives nouvelles des linguistiques de corpus portant sur des corpus textuels. Cette réflexion autour de ce type de corpus est d'autant plus importante à l'heure du numérique qu'on a affaire à une conception nouvelle d'un objet ancien, qui implique de devoir poser un regard neuf sur un objet pourtant familier. La prise de conscience du biais pouvant être introduit par cette habitude du corpus textuel doit encourager à mener une réflexion constamment renouvelée sur la place du chercheur et la vision qu'il donne des phénomènes. C'est sur ce type de questions que se rejoignent « nouveaux » corpus textuels et corpus vidéo comme le montrent les articles de la deuxième partie de ce numéro.

### **3. Spécificités de la constitution et du traitement de corpus vidéo**

La nature du corpus vidéo, à la différence des corpus textuels, rend plus saillante encore la question de la place du chercheur dans les analyses de corpus. Si les corpus textuels peuvent être variés, les données qu'ils contiennent ne nécessitent pas toujours l'intervention d'un intermédiaire et peuvent être considérées comme des données « brutes », que le chercheur peut, s'il le souhaite, directement observer et conserver en l'état. Dans le cas de données vidéo, l'accès aux données primaires est impossible. Avant même de constituer un corpus, la collecte des données nécessite toujours l'intervention d'un œil autre que celui du chercheur : celui de la caméra. Partant de là, différentes questions émergent, inhérentes à ce nouvel intermédiaire.

Les dispositifs d'enregistrement constituent en effet un facteur à prendre en compte à plusieurs titres. D'abord, parce que l'œil de la caméra n'est pas l'œil humain, et que plusieurs dispositifs sont nécessaires pour rendre compte de ce que voit l'observateur. De cela, découle le fait que les données recueillies constituent au final la vision que le chercheur a des événements, et restreint de ce fait considérablement le champ de recherche. Cet élément doit par conséquent être pris en compte par le chercheur au moment de la mise en place de son dispositif. Ensuite, parce que le dispositif d'enregistrement est visible et que sa présence peut modifier les réalisations produites par les locuteurs observés. On doit donc se demander dans quelle mesure on peut encore parler de données « authentiques », ne sachant pas ce



qu'elles auraient été sans la présence de ce dispositif. Or, ce point pose problème sur le plan théorique dès lors qu'on garde à l'esprit que les linguistiques de corpus se réclament d'une analyse des usages réels de la langue. Enfin, parce que la temporalité des événements joue un rôle central dans la collecte de ce type de données. Au contraire des données textuelles qui sont disponibles en tout temps (ou presque), les données vidéo ne sont disponibles qu'une fois qu'elles ont été collectées. Si un phénomène échappe au dispositif d'enregistrement, il ne pourra plus jamais être collecté dans les mêmes conditions, au même moment. Il n'y a pas de retour en arrière possible lors de la collecte, ce qui implique une réflexion préalable au recueil d'informations encore plus poussée que dans le cas des corpus textuels. Les trois articles qui suivent mettent en évidence, chacun dans leur domaine, ces enjeux spécifiques aux corpus vidéo.

Les deux premiers articles, bien que traitant de domaines relativement différents, se rejoignent sur l'ensemble des enjeux mis en évidence dans le cadre de la collecte de données vidéo. Tous deux mettent en avant la dimension pratique des problèmes posés par la collecte de données vidéo, que ce soit dans le cadre de l'enregistrement de réunions de travail pour Vassiliki Markaki ou dans le but de collecter des données authentiques sur les interactions entre apprentis-serveurs et clients d'un restaurant expérimental dans le cas de Clémentine Hugol-Gential. Les deux auteurs exposent les différentes contraintes spatiales et matérielles auxquelles peut être confronté le chercheur lors de l'installation de son dispositif d'enregistrement et mettent en évidence leurs conséquences méthodologiques et épistémologiques non seulement sur la collecte des données mais aussi sur l'analyse des phénomènes. Certaines questions émergent cependant de façon quelque peu différente. Vassiliki Markaki montre par exemple que la temporalité inhérente aux interactions en cours et la subjectivité du chercheur interprétant les événements qui se déroulent sous ses yeux le conduisent souvent à repenser l'agencement de son dispositif et à orienter différemment les objectifs de son analyse au moment de la collecte des données. Clémentine Hugol-Gential, quant à elle, relève un problème inhérent à la présence des dispositifs d'enregistrement et pose la question de son influence sur les usagers et sur leurs réalisations discursives ou leur comportement.

Dans son article, Isabel Colon de Carvajal s'inscrit dans une perspective similaire à celle des deux articles précédents puisque le corpus qu'elle analyse est constitué en vue d'analyser les pratiques des professionnels des centres d'appels en situation naturelle. Il est intéressant de voir que son travail met en évidence le fait que les objectifs d'analyse sont ici aussi constamment repensés au cours même de la collecte des données. Ce qui semble transparaître de façon plus singulière cependant, c'est le fait que le dispositif d'enregistrement, justement parce qu'il est un intermédiaire entre le chercheur et les données, ne constitue pas qu'un biais éloignant le chercheur, mais au contraire peut être à même de rendre compte de points de vue complémentaires qu'il n'aurait été possible d'obtenir au même moment sans y avoir recours. Paradoxalement, le dispositif d'enregistrement qui semblait à l'origine limiter le chercheur dans l'accès aux données lui permet ici d'obtenir de nouvelles informations et d'observer des phénomènes ne pouvant être directement appréhendés.

Le corpus vidéo nous semble donc bien être emblématique du corpus en tant qu'ensemble construit en perpétuel mouvement et se prête donc particulièrement bien aux approches *bottom-up* revendiquées par les linguistiques de corpus. On remarquera aussi que les corpus vidéo répondent généralement à un besoin de saisir un ensemble de réalisations spontanées, produites dans des situations naturelles, alors même que le recours nécessaire aux dispositifs d'enregistrement semble biaiser quelque peu l'authenticité des réalisations produites par les sujets observés. Il n'aura été question dans cette partie que des problèmes inhérents à la conception du corpus et à la collecte des données. La dernière partie entend aborder les questions soulevées par les dimensions temporelles, gestuelles et orales inhérentes à l'analyse des corpus multimodaux de leur constitution à leur exploitation.

#### **4. Les enjeux des linguistiques de corpus à la lumière de la multimodalité**

Trois articles apportent enfin un éclairage distinct et complémentaire sur les problématiques touchant à l'analyse de la multimodalité et au traitement de corpus hétérogènes.

À partir de l'analyse de l'enregistrement d'une visite guidée à laquelle participent trois personnes, Lorenza Mondada pose la question des

éléments observables dans ce type de corpus en vue d'une analyse des interactions. Ce genre d'analyse prend en compte des éléments aussi divers et hétérogènes que le langage (prosodie, énoncés), les regards, les gestes, les postures corporelles ou encore les mouvements des participants. De ce fait, la constitution et l'exploitation du corpus ne se fait plus dans un but unique mais rassemble simultanément différents objets d'analyse. De plus, au même titre que l'ont fait ressortir les articles portant sur les corpus vidéo, l'auteur met en évidence le fait que la tâche du chercheur est d'autant plus complexe que les phénomènes émergent au fur et à mesure de l'interaction et sont du même coup difficilement prévisibles. On est par conséquent confronté avec ce type de corpus à un ensemble de problèmes inhérents à la nature des données qui se construisent en temps réel. La temporalité des données multimodales et interactionnelles constitue donc un facteur qui doit nécessairement être traité de façon adaptée et implique la mise en place de stratégies spécifiques de la part du chercheur.

Le rôle du chercheur est au centre du travail présenté par Geneviève Pinard-Prevost au sujet des corpus audio d'enregistrement de conversations naturelles entre proches. Les linguistiques de corpus, on l'a vu, sont issues des linguistiques du TAL et ont souvent recours à des outils informatisés facilitant le traitement des données et permettant l'exploitation de corpus plus vastes. Si cette approche peut sembler favoriser non seulement une analyse de phénomènes plus variés mais aussi la mise en évidence de régularités appuyées sur un plus grand nombre d'occurrences, l'auteur montre ici que le recours à l'informatique ne permet pas de rendre compte de certains phénomènes particuliers comme les indices de prosodie qui interviennent souvent lors de chevauchements de parole ou de conversation parallèle. Plusieurs données sont alors contenues dans le signal vocal et les outils informatiques ne sont pas à même à l'heure actuelle de traiter ce type de données hétérogènes.

Enfin, Samira Moukrim présente quant à elle les difficultés liées à la constitution et à l'exploitation d'un corpus oral de deux langues à tradition orale : le berbère tamazight et l'arabe marocain. Si l'auteur ne se trouve pas ici confrontée à l'analyse de données multimodales, le fait qu'il s'agisse de deux langues à tradition orale pose le problème de l'hétérogénéité des données à exploiter. L'hétérogénéité des données se situe à deux niveaux : d'une part au niveau du recueil des données,

puisque l'on a affaire à des langues différentes, et d'autre part au niveau de leur transcription, puisque la codification graphique de chacune de ces langues présente de grandes disparités. Se pose alors, en plus des questions inhérentes à l'authenticité des données produites lors d'enregistrements audio et à la représentativité des informations recueillies, la question de l'unité du corpus. Si le corpus est un ensemble construit, on peut se demander dans quelle mesure on peut encore le considérer comme un ensemble lorsqu'il rassemble des données d'une telle hétérogénéité et si la finalité commune en vue de laquelle il a été constitué représente un critère suffisant.

### **Bilan et perspectives**

Les questions de la représentativité du corpus, de la place du chercheur et de la mutualisation des données constituent des problématiques communes à tout type de corpus. En dépit de cette transversalité mise en évidence dans le numéro 54, les différents travaux rassemblés dans le présent numéro nous conduisent inévitablement à ne plus parler « du » corpus, mais bien « des » corpus, à ne plus parler de « linguistique de corpus », mais plutôt « des linguistiques de corpus ». Les auteurs ont su mettre en lumière ici les spécificités inhérentes à chacun des types de corpus auxquels ils ont eu affaire. On a pu ainsi voir quels étaient les enjeux actuels des analyses de corpus textuels face à la révolution numérique, mais aussi saisir les problèmes posés par la mise en place des dispositifs d'enregistrement vidéo et l'analyse de données multimodales, tant du point de vue du chercheur que du point de vue des usages produits par les sujets observés.

Cependant, si certaines de ces spécificités sont communes à chaque nature de corpus, tous les travaux pointent un élément qui nous a semblé primordial dans la réflexion que nous entendions mener par le biais de ces deux numéros consacrés aux corpus. Il s'agit du fait que le chercheur, en linguistique de corpus, travaille sur un objet vivant, en constante évolution, et que bien souvent, des phénomènes apparaissent au cours du travail de recherche, que ce soit lors de la collecte des données comme lors de leur analyse. De ce fait, le chercheur doit mettre en place des stratégies adaptées non pas à un type de corpus, mais bien à son propre corpus en tant qu'objet unique et singulier. On peut donc se demander, pour clore cet état de la question à l'heure

où nous écrivons, s'il ne faudrait pas plutôt aller jusqu'à parler de « linguistiques des corpus », tant chaque méthode ou modèle d'analyse semble ne pouvoir s'appliquer de façon satisfaisante qu'à un seul et même corpus. Si cette perspective semble remettre en question la possibilité de modéliser certains phénomènes, ou du point de vue méthodologique, d'élaborer des modèles d'analyse applicables à plusieurs corpus, il n'en est rien. Cela nous semble au contraire montrer que si les réflexions autour des corpus sont monnaie courante, un long chemin reste encore à parcourir.

### Références bibliographiques

- HABERT B., NAZARENKO A. & SALEM A.,  
1997, *Les linguistiques de corpus*, Paris, Armand Colin.
- Langages* 2012/3 (n° 187), GARRIC N. & LONGHI J., L'analyse de corpus face à l'hétérogénéité des données.
- Langages* 2008/3 (n° 171), CORI M. *et al.* (éd.), Construction des faits en linguistique : la place des corpus.
- MAYAFFRE D., 2005, « Rôle et place des corpus en linguistique : réflexions introductives », *Texte !* [en ligne], décembre 2005, vol. X, n° 4. Disponible sur : [www.revue-texto.net/Reperes/Themes/Mayaffre\\_Corpus.html](http://www.revue-texto.net/Reperes/Themes/Mayaffre_Corpus.html). (Consultée le 20 octobre 2012).
- Revue française de linguistique appliquée*,  
2007/1 (vol. XII), *Corpus : bilans et perspectives*.
- TEUBERT W., 2009, « Corpus Linguistics : An Alternative », *Critical Discourse Analysis I. Les notions de contexte et d'acteurs sociaux*, *Semen* 27, version en ligne <http://semen.revues.org/8912>. Consultée le 19/10/2012.
- WILLIAMS G., 2005, *La linguistique de corpus*, Rennes, Presses universitaires de Rennes.